The AAPS Journal 2004; 6 (3) Article 21 (http://www.aapsj.org).

Transporter and Ion Channel Gene Expression After Caco-2 Cell Differentiation Using 2 Different Microarray Technologies

Submitted: July 9, 2003; Accepted: May 6, 2004; Published: September 7, 2004.

Christopher P. Landowski,[1] Pascale Anderle,[2,3,4] Duxin Sun,[1,5] Wolfgang Sadee,[2,3] and Gordon L. Amidon[1]

[1]Department of Pharmaceutical Sciences, College of Pharmacy, University of Michigan, Ann Arbor, MI 48109
[2]Department of Biopharmaceutical Sciences, School of Pharmacy, University of San Francisco, San Francisco, CA 94143
[3]Department of Pharmacology, College of Medicine and Public Health, Ohio State University, Columbus, OH 43210
[4]Current address: Swiss Institute for Experimental Cancer Research (ISREC), 1066 Epalinges s/Lausanne, Switzerland
[5]Current address: Department of Pharmaceutics, College of Pharmacy, Ohio State University, Columbus, OH 43210

ABSTRACT

mRNA expression profiles had previously been measured in Caco-2 cells (human colonic carcinoma cells) using either custom-designed spotted oligonucleotide arrays or Affymetrix GeneChip oligonucleotide arrays. The Caco-2 cells used were from different clones and were examined under slightly different culture conditions commonly encountered when Caco-2 cells are used as a model tissue for studying intestinal transport and metabolism in different laboratories. In this study, we compared gene expression profiles of Caco-2 cells generated with different arrays to assess the validity of conclusions derived from the 2 independent studies, with a focus on changes in transporter and ion channel mRNA expression levels on Caco-2 cell differentiation. Significant changes in expression levels upon differentiation were observed with 78 genes, with probes common to both arrays. Of these, 18 genes were upregulated and 36 genes were downregulated. The 2 arrays yielded discrepant results for 24 genes, showing significant changes upon differentiation. The results from the 2 arrays correlated well for genes expressed above average levels ($r = 0.75$, $P < 0.01$, n = 25) and poorly for genes expressed at low levels ($r = 0.08$, $P > 0.05$, n = 25). Overall correlation across the 2 platforms was $r = 0.45$ ($P < 0.01$) for the 78 genes, with similar results from both arrays. Despite differences in experimental conditions and array technology, similar results were obtained for most genes.

KEYWORDS: microarrays, Caco-2 cells, transporter, ion channel, platform comparison.

INTRODUCTION

The Human Genome Project has uncovered over 30 000 gene sequences that comprise the entire human transcriptome.[1] Currently, the functions of many genes remain unknown, but high-throughput analysis of mRNA expression can provide clues to their potential functions.[2] Global snapshots of gene expression in tissues or in cell lines under different conditions can help identify candidate genes that may be involved in a variety of cellular processes.[3] In one such application, carrier-mediated drug permeability was compared between Caco-2 cells in vitro and in human intestines in vivo, on the one hand, and transporter gene expression on the other.[4] Differential expression of specific genes in normal and diseased tissues may lead to diagnostic genetic markers or to targets for treatment strategies.[5] Global gene expression analysis has also resulted in the identification of transporters that play a role in the oral absorption of valacyclovir in humans.[6]

The most common tools for performing high-throughput expression measurements include complementary DNA (cDNA) microarrays, oligonucleotide microarrays, or serial analysis of gene expression (SAGE). These techniques allow the simultaneous determination of mRNA expression from tens of thousands of genes. The Affymetrix GeneChip (Affymetrix Inc, Santa Clara, CA) is a widely used commercial array with short oligonucleotide gene probes photo-lithographically synthesized on a silicon wafer.[7] Oligonucleotide arrays can also be prepared by spotting presynthesized oligonucleotide probes (50 to 70 mers) onto chemically treated glass slides.[8,9] Microarrays can also be obtained by spotting cDNA probes onto a slide.[10] The latter provides for good sensitivity but may suffer from lower specificity if closely related genes are measured, owing to potential cross-hybridizations.

There is no established standard methodology in the research community for the performance of global expression studies. In addition, microarray and SAGE expression data for a variety of systems are freely available on Internet Web sites.[11] Therefore, it is not uncommon to have several sets of expression data derived from various detection methods for a given system. Critical comparisons would allow the consolidation of data from many different platforms and strengthen the confidence in the expression analyses process.

Several studies on comparisons of array platforms have been published.[12-18] Expression analysis with spotted microarrays is not as easy to compare with Affymetrix GeneChip or SAGE analysis methods because the spotted microarray expression

Corresponding Author: Gordon L. Amidon, Department of Pharmaceutical Sciences, College of Pharmacy, University of Michigan, Ann Arbor, MI 48109. Tel: 734-764-2464. Fax: 734-763-6423. Email: glamidon@umich.edu.

data are usually reported as a ratio of the signal from a control to a signal from an experimental sample. However, expression levels relative to an identical reference would allow comparisons between the 2 methods. Indeed, such cross-platform correlations between these 2 array systems using NCI-60 cancer cell lines were attempted, but the measurements from the 2 platforms exhibited poor correlation. The authors concluded that probe-specific factors such as G-C content, sequence length, average signal intensity, and cross-hybridization influenced the measurements in the 2 platforms differently, resulting in poor correlations between them.[15] In our experience, the method of data normalization also strongly affects correlations. Tan et al described the evaluation of 3 commercial microarray platforms using a standardized input RNA sample.[17] The authors compared 2 commercial 25-mer and 30-mer oligonucleotide arrays (Affymetrix and Amersham [Amersham Biosciences, Piscataway, NJ], respectively) with a cDNA array from Agilent (Agilent Technologies, Palo Alto, CA). The authors observed only modest cross-platform correlations (Pearson's linear correlation ranging from 0.48 to 0.60) and suggest that these differences may be related to the intrinsic properties of the microarrays themselves. Recently, Barczak et al reported a comparison of gene expression measurements made with identical RNA preparations using Affymetrix 25-mer oligonucleotide arrays and 2 commercially available collections of ~70-mer spotted long nucleotide arrays.[18] The authors evaluated the agreement between measurements of 7344 common genes and in contrast to previous reports[14,15,17] found very good overall agreement between the 3 arrays examined (Pearson's linear correlations ranging from 0.80 to 0.83). In common with previous observations,[17,19] the authors found that these correlations improved (Pearson's correlation 0.86 to 0.89) when low-intensity signals on either array type were excluded.

In this report, we describe the comparison of the changes in gene expression profiles upon differentiation of Caco-2 cells reported earlier using 2 different methods.[4,8] Thus, Sun et al[4] reported Caco-2 gene expression profiles obtained using the Affymetrix GeneChip microarray, whereas Anderle et al[8] reported changes in transporter and ion channel mRNA expression levels on Caco-2 cell differentiation obtained using a custom-designed, spotted 70-oligomer oligonucleotide microarray.[8] The genes examined are involved in the intestinal transport of a variety of therapeutically important entities.[8,20] The major objective of this report was to compare mRNA expression in Caco-2 cells for genes encoding transporters and ion channels measured in the 2 studies. A second objective was to evaluate measured differences in expression changes upon differentiation of Caco-2 cells obtained using different clones and slightly different culturing conditions.

Materials and Methods

Materials

All reagents, cell cultures, and cell culture accessories used in the study have been described elsewhere.[4,8] The Affymetrix GeneChip oligonucleotide array data reported earlier[4] was obtained using the Affymetrix GeneChip U95A and was purchased from Affymetrix Inc. The preparation of the custom-designed spotted oligonucleotide array used to obtain expression data for transporter and ion channel genes has been described previously.[8]

Preparation of mRNA, Labeling, Hybridization, and Scanning

Custom-Designed Spotted Array

Caco-2 cells were cultured as described in Anderle et al.[8] Briefly, Caco-2 cells were cultured at 37°C, 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/L D-glucose, 20% fetal bovine serum, nonessential amino acids (1% vol/vol), 100 U/mL penicillin, and 100 g/mL streptomycin. Caco-2 total RNA samples were prepared as described in Anderle et al.[8] Messenger RNA was extracted from cells using the FastTrack 2.0 Kit (Invitrogen, Carlsbad, CA). Samples of 2 µg mRNA or 20 µg total RNA, respectively, were labeled with either Cy3 Dye or Cy5 Dye by amino-allyl coupling (http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf), resuspended in 20 µL HEPES buffer (25 mM, pH 7.0) containing 1 µL of total RNA, 1.5 µL of polyA$^+$ 0.45 µL of 10% SDS, and hybridized to the slides for 16 hours at 65°C. Slides were washed and dried before scanning, according to DeRisi et al.[21] Slides were scanned on a GenePix 4000A (Axon Instruments, Union City, CA) to detect Cy3 and Cy5 fluorescence. Housekeeping genes and negative controls were the same as in the Atlas 1.2 Human Array by Clontech (Palo Alto, CA).

Affymetrix Oligonucleotide Array

Caco-2 cells were cultured at 37°C, 5% CO_2 in DMEM supplemented with 4.5 g/L D-glucose, 10% fetal bovine serum, nonessential amino acids (1% vol/vol), 100 U/mL penicillin, and 100 g/mL streptomycin.[4] Caco-2 total RNA samples were prepared as described in Sun et al.[4] Caco-2 cells were homogenized in TRIzol reagent (Invitrogen, Carlsbad, CA), and the RNA phase was separated by chloroform; total RNA was precipitated with isopropyl alcohol and washed with 80% ethanol. Total RNA was further cleaned with RNeasy mini kit (Qiagen, Valencia, CA). First-strand cDNA was transcribed from total RNA using T7-$(dT)_{24}$ oligomer primer and SSII reverse transcriptase at 42°C. The second strand cDNA was synthesized from first-strand cDNA using DNA ligase, DNA polymerase I, and T4 DNA polymerase at 16°C

(SuperScript Choice System for cDNA synthesis kit, Gibco), then cleaned with phase-locking gel. Biotin-labeled cRNA synthesized from the double strand cDNA using T7 RNA polymerase-catalyzed in vitro transcription in the presence of biotin-labeled nucleoside triphosphate (NTP) (BioArray high-yield RNA transcription labeling kit, Enzo Biochem, Farmingdale, NY), then fragmented at 95°C. Biotin-labeled cRNA was heated at 99°C for 5 minutes in hybridization cocktail including hybridization controls (*Bio B*, *C*, *D*, and *Cre*) and hybridized with GeneChip (Affymetrix) at 42°C for 16 hours. The GeneChip was then washed with nonstringent wash buffer at 50°C and stained with streptavidin phycoerythrin (SAPE) solution. After washing at 25°C, the GeneChip was scanned with a laser scanner (Affymetrix). The gene expression profiles were analyzed by Affymetrix Microarray Suite and Data Mining Tool software. Sample quality was assessed by 5'/3' ratios of endogenous controls (GAPDH, beta-actin, transferrin receptor, and ISGF-3), and exogenously added RNA (Bio B, C, D, and Cre).

Gene Expression Data

The Affymetrix GeneChip[4] and custom-designed spotted oligonucleotide[8] Caco-2 gene expression data, previously reported in literature, were used in this comparison study. The Affymetrix U95A microarrays contained 12 599 probe sets, while the spotted array contained 750 oligonucleotide probes. All probe sequences, the 12 599 Affymetrix probes, and 632 70-oligomer spotted array probes were mapped to UniGene clusters in order to match probe sequences. Only UniGene clusters that were represented on both platforms were considered for the comparison. A total of 115 genes were found to be common between the 2 platforms before imposing cutoff values.

The spotted array data represent the mean of 12 probes (3 slides, with 4 gene probes per slide). The gene expression data obtained with the spotted array data were further refined by locally weighted linear regression curve fit (Lowess) normalization because the data were not normally distributed and because Lowess normalization introduces less error in analyzing such data sets.[22] The raw data for the custom-designed spotted oligonucleotide and the Affymetrix array are presented as supplemental information (available in online version). The custom array data is also available as accession number GSE1368 at http://www.ncbi.nlm.nih.gov/geo/. The Affymetrix array data represent the mean of 32 probes (2 arrays, 16 gene probes per array). The Affymetrix data[4] were normalized by the default method performed by the Affymetrix Microarray Suite and Data Mining software. The global normalization analysis uses a total across chip average to normalize the data and is considered to be appropriate for normalization of large data sets such as the one generated by the Affymetrix GeneChip. The normalized data from each platform were used to determine the changes in various gene expression levels. The change in expression levels of a gene in each platform was expressed as a Log to the base 2 ($\log_2$) ratio of expression levels after and before Caco-2 cell differentiation. Log to the base 2 ratios are customarily used in order to compress the data set values but, at the same time, allow any differences to be clearly distinguished.

Correlation Computations

SPSS (version 9.0, SPSS Inc, Chicago, IL) was used to calculate Pearson linear correlation coefficients and Spearman rank-order correlation coefficients for the matched expression measurements. GraphPad Prism 3.0 (GraphPad, San Diego, CA) was used to perform *t* tests and analysis of variance (ANOVA) analysis.

RESULTS

Probe Sequence Overlap

Of the 115 transporter and ion channel genes present on both the spotted array and the GeneChip, the majority of the gene probe sequences were different between the 2 platforms. Thus, only 19 of the 70-mer probe sequences in the spotted array exhibited some overlap with the 16 25-mer probe sequences of the Affymetrix GeneChip array. When there was overlap, on average the matching 70-mer probe coincided with one of the 16 25-mer sequences on the GeneChip array. The maximum overlap of a matching 70-mer spotted array probe was with 10 of the 16 25-mer probes of the GeneChip array, which occurred for the arsenite translocating ATPase (ASNA1) gene. The Affymetrix and custom arrays detected comparable downregulation for the ASNA1 gene, but the respective $\log_2$ ratios (-0.12 and -0.21) may not be significant. Of the 19 70-mer probe sequences in the spotted array that exhibited overlap with the GeneChip array, 11 were expressed at high levels in both platforms. The $\log_2$ ratio values generated for this group of 11 matching genes were found to correlate very well (Pearson value $r = 0.75$, $n = 11$).

Changes in Transporter and Ion Channel Gene Expression Levels Upon Caco-2 Cell Differentiation

Before comparing expression changes measured with the Affymetrix GeneChip array and a custom-designed spotted array, it is essential to determine the reliability and significance of the changes in mRNA expression levels in the 2 methods. The error associated with measuring relative mRNA levels between 2 identical poly(A+) samples obtained from Caco-2 cells cultured for 5 days and labeled with Cy3 or with Cy5 has previously been reported for the custom-designed oligonucleotide array.[8] When the mean fluorescence intensities from 4 observations per probe (spotted

Table 1. Genes Exhibiting Increased Expression Levels in Both Arrays After Caco-2 Differentiation: GeneChip $\log_2$ (16 day/4 day) and Microarray $\log_2$ (14 day/5 day)

Unigene	Description	GeneChip*	Microarray*
Hs.79428	BCL2/adenovirus E1B 19kDa interacting protein 3	**2.59**	**0.65**
Hs.90786	ATP binding cassette, sub-family C (CFTR/MRP), member 3	**2.41**	**0.48**
Hs.2257	Vitronectin (serum spreading factor, somatomedin B, complement S-protein)	**1.83**	**1.15**
Hs.89436	Cadherin 17, LI cadherin (liver-intestine)	**1.62**	**0.18**
Hs.50868	Solute carrier family 22 (organic cation transporter), member 1-like	**1.24**	**0.13**
Hs.166196	Cytokeratin 2	**1.21**	**0.03**
Hs.234642	Aquaporin 3	**1.00**	**0.19**
Hs.553	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	**0.98**	**0.24**
Hs.184276	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1	**0.87**	**0.05**
Hs.24030	Solute carrier family 31 (copper transporters), member 2	**0.77**	**0.11**
Hs.125856	ATP binding cassette, sub-family B (MDR/TAP), member 7	**0.73**	**0.11**
Hs.194693	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	**0.70**	**0.05**
Hs.159322	Solute carrier family 35 (UDP-N-acetylglucosamine transporter), member A3	**0.67**	**0.20**
Hs.119529	Apolipoprotein (a) related gene C	**0.63**	**0.01**
Hs.77239	Solute carrier family 22 (organic cation transporter), member 4	**0.58**	**0.13**
Hs.663	Cystic fibrosis transmembrane conductance regulator, ATP binding cassette (sub-family C, member 7)	**0.53**	**0.02**
Hs.111894	Lysosomal-associated protein transmembrane 4 alpha	**0.48**	**0.08**
Hs.76460	Solute carrier family 38, member 3	**0.47**	**0.07**

*Changes less than a ratio of 0.45 may not be significant.

4 times on a single slide) were plotted, the deviations from the line of unity were 1.5-fold maximally. Using 3 replicate slides (ie, 12 total observations), a 1.3- to 1.5-fold change in mRNA expression level over the control could be reliably measured, depending on the signal intensity. These results suggested that a 0.45 $\log_2$ ratio cutoff value would be appropriate.[8,23-25] The error associated with the Affymetrix GeneChip measurements was determined in a similar manner. Plotting normalized intensity values from 12 559 probe sets obtained from identical 4-day Caco-2 cell mRNA samples indicated that deviations from the line of unity ranged from 1.7-fold to 1.1-fold, depending on the signal intensity. Thus, a 0.45 (or 1.4-fold change) $\log_2$ cutoff value also appears to be appropriate for Affymetrix arrays.[26] Log_2 ratios of differentiated to undifferentiated Caco-2 expression data from the 115 matching genes were screened using a cutoff value of 0.45. In addition, genes exhibiting expression changes that met the cutoff value criterion in 1 array but not in the other were also retained for comparison; however, the data from the other array may not represent a statistically significant change in expression. In comparing measurements collected with different array types, such data retention based on high differential expression in at least 1 array has been reported earlier.[18] A total of 78 genes out of the 115 matching genes satisfied these requirements. Eleven genes in both data sets satisfied the 0.45 cutoff criterion.

Expression changes following Caco-2 cell differentiation for these 78 genes are listed in Tables 1, 2, and 3. Both arrays determined that 18 genes were upregulated and 36 genes were downregulated. Discrepant changes in expression levels were observed for the 24 remaining genes (Table 3). Several transporter and ion channel genes exhibited significant changes in expression upon Caco-2 differentiation. The Affymetrix array determined that the BCL2/adenovirus E1B 19-kDa interacting protein 3 was most upregulated after differentiation as indicated by a $\log_2$ (16-day/4-day) value of 2.59. In the spotted microarray data, vitronectin was the most upregulated gene, with a $\log_2$ (14-day/5-day) ratio of 1.15. The most downregulated gene as determined by GeneChip analysis was the solute carrier family 4, sodium bicarbonate cotransporter, member 8, with a $\log_2$ (16-day/4-day) value of -2.24. The most downregulated gene as determined by spotted array was the solute carrier family 7 (cationic amino acid transporter, y+ system), member 5, which had a $\log_2$ ratio of -1.20. The $\log_2$ ratios generated by GeneChip were in the range of +2.59 to -2.24, while the spotted array expression differences were in the range of +1.15 to -1.20.

Expression Correlations Between Microarrays

The intraplatform correlation of normalized intensities for 12 559 probe sets with the Affymetrix array was satisfactory (r = 0.94-0.96) and similar to that reported earlier.[18,27,28] For the custom-design spotted microarray, the correlation of gene expression measurements from 3 replicates exhibited correlation coefficients ranging from 0.94 to 0.97, similar to that reported for spotted long oligonucleotide microarrays.[18,25,29] Further, the Affymetrix 4-day sample replicates were similar

Table 2. Genes Exhibiting Decreased Expression Levels in Both Arrays After Caco-2 Differentiation: GeneChip $\log_2$ (16 day/4 day) and Microarray $\log_2$ (14 day/5 day)

Unigene	Description	GeneChip*	Microarray*
Hs.132136	Solute carrier family 4, sodium bicarbonate cotransporter, member 8	**-2.24**	**-0.04**
Hs.159557	Karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	**-2.11**	**-0.64**
Hs.139336	ATP binding cassette, sub-family C (CFTR/MRP), member 4	**-1.79**	**-0.08**
Hs.184601	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	**-1.66**	**-1.20**
Hs.952	Solute carrier family 10 (sodium/bile acid cotransporter family), member 1	**-1.49**	**-0.07**
Hs.22891	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 8	**-1.47**	**-0.18**
Hs.24040	Potassium channel, subfamily K, member 3	**-1.46**	**-0.18**
Hs.89512	ATPase, Ca^{2+} transporting, plasma membrane 2	**-1.37**	**-0.01**
Hs.75379	Solute carrier family 1 (glial high affinity glutamate transporter), member 3	**-1.22**	**-0.49**
Hs.111967	Solute carrier family 30 (zinc transporter), member 3	**-1.16**	**-0.13**
Hs.110736	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2	**-1.11**	**-0.48**
Hs.108660	ATP binding cassette, sub-family C (CFTR/MRP), member 5	**-1.11**	**-0.02**
Hs.101408	Branched chain aminotransferase 2, mitochondrial	**-1.10**	**-0.05**
Hs.84190	Solute carrier family 19 (folate transporter), member 1	**-1.07**	**-0.23**
Hs.89433	ATP binding cassette, sub-family C (CFTR/MRP), member 1	**-1.06**	**-0.16**
Hs.1526	ATPase, Ca^{2+} transporting, cardiac muscle, slow twitch 2	**-0.94**	**-0.45**
Hs.77572	BCL2/adenovirus E1B 19kDa interacting protein 1	**-0.91**	**-0.10**
Hs.129683	Ubiquitin-conjugating enzyme E2D 1 (UBC4/5 homolog, yeast)	**-0.91**	**-0.02**
Hs.7594	Solute carrier family 2 (facilitated glucose transporter), member 3	**-0.81**	**-0.94**
Hs.23965	Solute carrier family 22 (organic anion transporter), member 6	**-0.80**	**-0.13**
Hs.121495	Potassium voltage-gated channel, Isk-related family, member 1	**-0.75**	**-0.11**
Hs.80658	Uncoupling protein 2 (mitochondrial, proton carrier)	**-0.69**	**-0.15**
Hs.54470	ATP binding cassette, sub-family C (CFTR/MRP), member 8	**-0.69**	**-0.03**
Hs.75231	Solute carrier family 16 (monocarboxylic acid transporters), member 1	**-0.68**	**-0.15**
Hs.25450	Solute carrier family 29 (nucleoside transporters), member 1	**-0.63**	**-0.11**
Hs.76781	ATP binding cassette, sub-family D (ALD), member 3	**-0.62**	**-0.69**
Hs.84974	Nuclear transcription factor Y, beta	**-0.61**	**-0.29**
Hs.12627	ATPase, H^+ transporting, lysosomal V0 subunit a isoform 2	**-0.57**	**-0.11**
Hs.153985	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 2	**-0.56**	**-0.19**
Hs.183556	Solute carrier family 1 (neutral amino acid transporter), member 5	**-0.55**	**-0.29**
Hs.3886	Karyopherin alpha 3 (importin alpha 4)	**-0.55**	**-0.22**
Hs.3112	Sodium channel, nonvoltage-gated 1, gamma	**-0.52**	**-0.01**
Hs.121499	Solute carrier family 6 (neurotransmitter transporter, glycine), member 9	**-0.51**	**-0.14**
Hs.79748	Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	**-0.50**	**-0.51**
Hs.76941	ATPase, Na^+/K+ transporting, beta 3 polypeptide	**-0.49**	**-0.07**
Hs.20716	Translocase of inner mitochondrial membrane 17 homolog A (yeast)	**-0.19**	**-0.55**

*Changes greater than a ratio of -0.45 may not be significant.

($P = 0.92$), while there was more variation in the 16-day replicates ($P < 0.05$) (Student t tests). Custom array 5-day ($P = 0.41$) or 14-day replicates ($P = 0.69$) were similar, as suggested by ANOVA. In both arrays, average expression levels in differentiated cells were significantly different from levels in undifferentiated cells ($P < 0.05$).

Log_2 (14-day/5-day) Caco-2 expression changes using the custom-designed spotted microarray were compared statistically with $\log_2$ (16-day/4-day) Caco-2 expression changes determined with Affymetrix array for the 78 genes by calculating both Pearson and Spearman coefficients. Overall, the correlation between the 2 methods for the entire set of 78 genes gave a Pearson correlation coefficient of 0.45 and a Spearman correlation coefficient of 0.48, both being statistically significant ($P < 0.01$) (Table 4).

The relatively low overall correlation may in part be related to the wide range in expression levels observed with the 78 gene probes in each array method. The expression level intensities exhibited a 450-fold difference between the highest and lowest expressed genes in Affymetrix GeneChip

Table 3. Genes with expression changes not in agreement after Caco-2 differentiation: GeneChip $\log_2$ (16 day/4 day) and Microarray $\log_2$ (14 day/5 day)

Unigene	Description	GeneChip*	Microarray*
Hs.33084	Solute carrier family 2 (facilitated glucose/fructose transporter), member 5	**1.36**	**-0.08**
Hs.107911	ATP binding cassette, sub-family B (MDR/TAP), member 6	**1.13**	**-0.14**
Hs.101813	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 2	**0.91**	**-0.01**
Hs.85838	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C	**0.82**	**-0.12**
Hs.123123	Chloride channel Ka	**0.81**	**-0.03**
Hs.101337	Uncoupling protein 3 (mitochondrial, proton carrier)	**0.77**	**-0.19**
Hs.111024	Solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	**0.52**	**-0.17**
Hs.78546	ATPase, Ca^{2+} transporting, plasma membrane 1	**0.51**	**-0.04**
Hs.106778	ATPase, Ca^{2+} transporting, type 2C, member 1	**0.50**	**-0.24**
Hs.157145	Tetracycline transporter-like protein	**0.48**	**-0.04**
Hs.76918	Niemann-Pick disease, type C1	**0.47**	**-0.01**
Hs.78629	ATPase, Na^+/K+ transporting, beta 1 polypeptide	**-0.10**	**0.68**
Hs.91139	Solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	**-0.79**	**0.16**
Hs.94395	ATP binding cassette, sub-family D (ALD), member 4	**-0.80**	**0.15**
Hs.176658	Aquaporin 8	**-0.86**	**0.20**
Hs.100001	Solute carrier family 17 (sodium phosphate), member 1	**-0.91**	**0.17**
Hs.158322	Solute carrier family 18 (vesicular monoamine), member 1	**-0.95**	**0.26**
Hs.211562	ATP binding cassette, sub-family A (ABC1), member 1	**-1.02**	**0.66**
Hs.6517	Amiloride-sensitive cation channel 1, neuronal (degenerin)	**-1.44**	**0.11**
Hs.2928	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	**-1.45**	**0.05**
Hs.38095	ATP binding cassette, sub-family A (ABC1), member 8	**-1.57**	**0.14**
Hs.187958	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8	**-1.67**	**0.12**
Hs.239106	Neutral and basic amino acid transport protein rBAT B(0,+)-type	**-2.26**	**0.19**
Hs.123639	Solute carrier family 6 (neurotransmitter transporter, GABA), member 11	**-0.45**	**0.24**

*Changes with absolute values less than 0.45 may not be significant.

Table 4. Comparison of Log_2 Ratios Sorted by Mean Expression Intensity

Genes in Correlation	Pearson Correlation (r)	Spearman Correlation (ρ)
Highest 25	0.75	0.77
Middle 25	0.49	0.34
Lowest 25	0.08	-0.14
All 78	0.45	0.32

measurements, while a 40-fold difference between the intensities of highest and lowest expressed genes was observed with spotted array measurements. Since the expression level range was quite large, the correlation between the 2 methods based on expression level intensity was examined. The mean intensity values from the 4-day Affymetrix GeneChip array data were used to sort the $\log_2$ ratios obtained with the 2 methods. The correlation between $\log_2$ ratios from the 2 methods was determined for expanding sets of genes with progressively decreasing average expression intensities. This was done for increasingly larger groups of genes, until the entire data set was compared. The results of the analysis are presented in Table 5 and Figure 1. It can be seen that the $\log_2$ ratio values for the 5 most highly expressed genes exhibit excellent correlation values (Pearson correlation 0.91, $P < 0.05$, and Spearman coefficient 0.80, $P < 0.05$). It is also clear from the results in Table 5 that the $\log_2$ correlation between the 2 methods decreases as the average gene expression intensity declines. Thus, the $\log_2$ ratio correlations decreased progressively from a 0.75 Pearson's value after comparing the top 25 most highly expressed genes (Figure 2A) to 0.49 for 25 genes with mid-level average expression (Figure 2B) to 0.08 for 25 genes with the lowest average expression (Figure 2C). Table 4 summarizes Pearson's correlation coefficients along with the Spearman ρ coefficients for the above gene sets.

Discussion

Several methods are available for large-scale gene expression analyses. Comparison of expression data derived from several different technologies may confer greater reliability to the results generated and minimize replication of costly microarray experiments. The present study describes a comparison of changes in transporter and ion channel gene

Table 5. Correlation of Log_2 Ratios Generated From the Affymetrix Array and the Custom Microarray as a Function of Average Expression Intensity (Average of Affymetrix and Custom Array Values)

Number of Genes	Pearson Correlation (r)	Spearman Correlation (ρ)	Mean Expression Intensity
5	0.91	0.80	10669
10	0.87	0.86	7192
15	0.83	0.84	5504
20	0.78	0.81	4490
25	0.75	0.77	3779
30	0.69	0.67	3258
35	0.67	0.67	2872
40	0.63	0.62	2568
45	0.62	0.61	2327
50	0.59	0.59	2127
55	0.51	0.46	1957
60	0.48	0.38	1811
65	0.51	0.41	1684
70	0.44	0.32	1571
75	0.43	0.33	1471
78	0.43	0.32	1415

expression levels after Caco-2 cell differentiation reported using a custom-designed spotted oligonucleotide microarray[8] with those reported using an Affymetrix GeneChip oligonucleotide microarray.[4] For 78 transporter and ion channel genes represented on both arrays, the correlation of data from the 2 methods was highly significant for genes that were expressed at above average levels. The correlations between the 2 methods became progressively poorer when low expression genes were included, which is consistent with previously observed low-intensity expression measurements.[15-19] These results are encouraging considering that the 2 studies being compared were conducted independently in 2 different labs using 2 different array technologies with Caco-2 cell lines that were different and grown under slightly different conditions.

Several factors may contribute to the observed disparity between the expression changes determined with the 2 different microarray systems. These factors include biological differences in the samples as well as differences in array design, in sample preparation methods, and in data normalization methods. Although the Caco-2 cell growth periods in the 2 independent studies were slightly different (4 days vs 5 days and 14 days vs 16 days), cells that are in culture 4 or 5 days are considered undifferentiated, while Caco-2 cells grown 14 to 16 days postseeding are considered to be differentiated.[30-34] Several reports confirm that gene expression changes minimally in Caco-2 cells over short time periods (1 or 2 days), compared with full differentiation.[23,24,26,35-37] Nevertheless, differences in Caco-2 clones and procedures used in the 2 laboratories may have contributed to differences in transporter and ion channel expression.

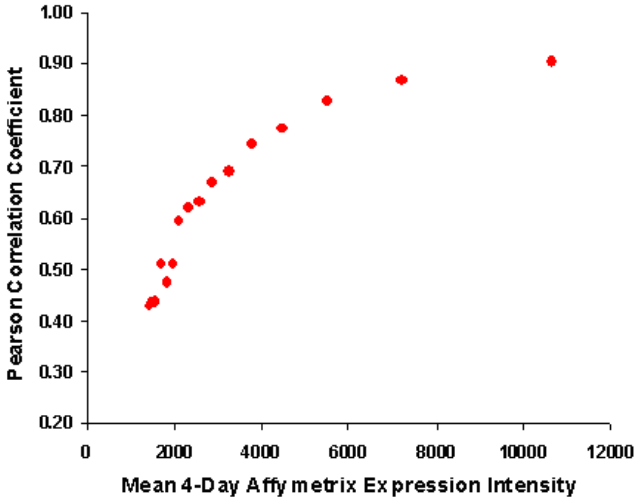

Figure 1. Pearson correlation of log_2 ratios as a function of mean expression intensity.

The correlation quality of the transporter and ion channel gene expression levels determined with the 2 arrays is expected to depend on similarity of probe-specific factors such as sequence overlap, sequence length, average signal intensity, and cross-hybridization in the 2 platforms. The probe sequences in the Affymetrix oligoarray are shorter than the probe sequences in the spotted array. Thus, the Affymetrix U95A chip employs 16 25-mer probes, while the spotted array contains 1 70-mer probe spotted 4 times on each slide. Further, the probes generated for the Affymetrix array are made to gene sequence from near the 3' end of the cDNA being assayed and are often made from sequence in the 3' untranslated region.[16] The probes for the spotted array were designed to be located as close to the 3' end of the coding sequence as possible.[8] Yet, of the 115 genes common to the 2 arrays, only 19 exhibited some degree of probe sequence overlap, while the remaining 96 genes did not overlap at all. On average, the 19 70-mer spotted array probes overlapped with 1 of the 25-mer Affymetrix probes; the maximum overlap observed was with 10 25-mer probes. The correlation of log_2 ratios generated from this group of overlapping probes for genes that were expressed above average levels was very good. This finding suggests that probe sequence overlap between the 2 different arrays may assure greater agreement of expression data generated by the 2 arrays. However, the reasonably good overall correlations obtained between the 2 methods for genes that exhibit no overlap at all indicates that probe sequence overlap is not a prerequisite for cross-platform agreement.

Differences in probe sequence lengths in 2 different arrays can also influence expression data generated by the arrays since the degree of nonspecific cross-hybridization is largely

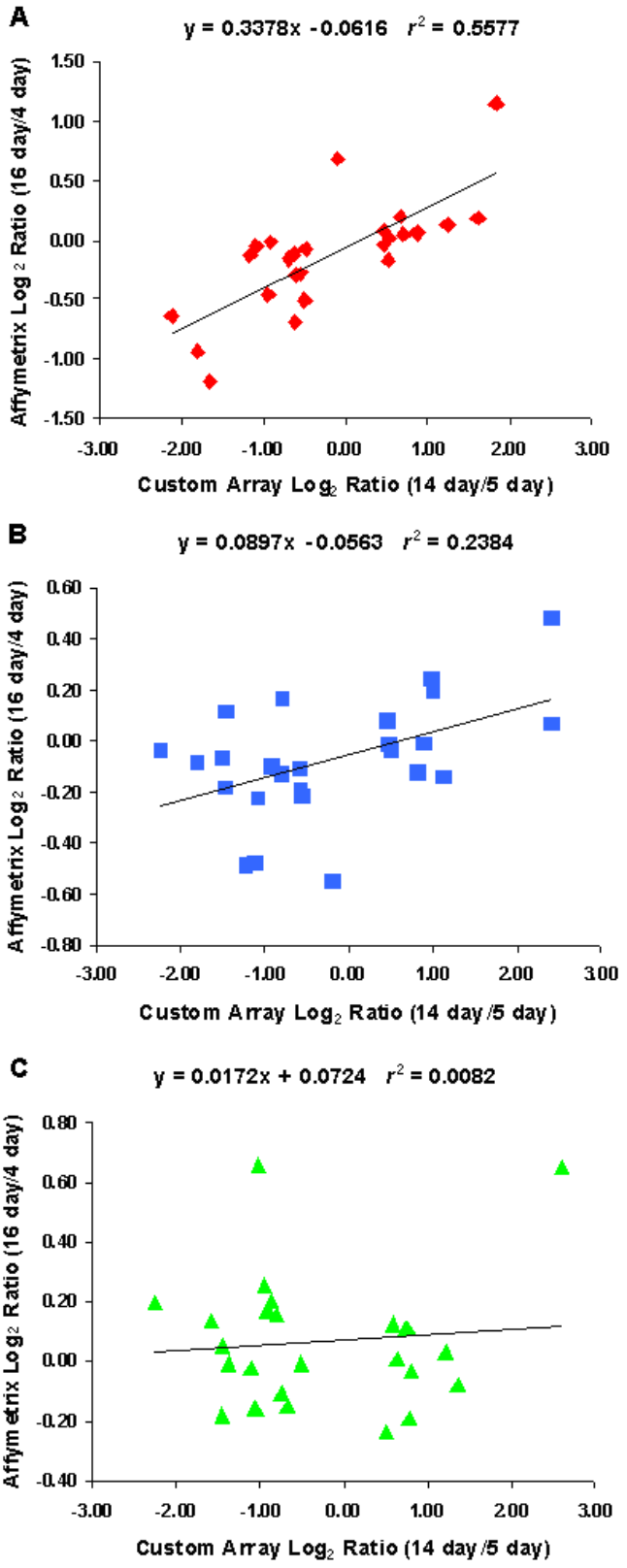

Figure 2. Pearson correlation of $\log_2$ ratios for 75 genes: (A) highest 25 genes, (B) middle 25 genes, and (C) lowest 25 genes.

dependent on sequence length. Thus, the longer 70-mer spotted array probe may exhibit higher potential for nonspecific hybridization, which could affect the background signal intensity observed even though the probes are selected for minimum cross-hybridizations. In order to control for this factor using the spotted microarray system, a separate reference sample and chip are necessary. Using a reference sample helps eliminate noise due to spot specific hybridization and signal detection.[10] The Affymetrix probes are smaller and have a built-in system for dealing with potential background noise through control probes. The control probe, called a mismatch probe (MM), is identical to the probe designed to be perfectly complementary to a target sequence, called the perfect match probe (PM), except for a single base mismatch in its center that allows direct subtraction of background and cross-hybridization signal. However, if strong hybridization to both PM and MM probe sets occurs, the MM signal masks the PM signal resulting in artificially low expression readout[38] and may skew the results for such genes. The large apparent fold changes associated with the Affymetrix data compared with the custom array data may indicate background hybridization of MM probes to certain genes.

Sample labeling methods are also important for producing consistent data. The sample preparation for hybridization to the Affymetrix microarray requires that biotin-labeled RNA be synthesized. The test sample RNA is converted into cDNA and then reconverted back into cRNA in order to incorporate biotinlyated bases. The biotin is required to attach the streptavidin phycoerythrin (SAPE) detection fluorophore. This method does amplify the messenger RNA signal, which may explain the larger range in signal seen in the Affymetrix intensity data. Affymetrix recommends using RNA because RNA/DNA hybrids are typically stronger than DNA/DNA hybrids. The sample preparation method for the spotted microarray is simpler as it yields labeled cDNA derived from mRNA. This labeling method also may be prone to less experimental error since the labeled material is being manipulated less. The labeling methods in both arrays do have in common the fact that the first strand of cDNA is made from the 3' poly-A tail region of the mRNA.

A total of 24 genes (31%) were found to be regulated differently in the 2 methods (Table 3). The observation that all the genes in this set were expressed at below average intensity levels suggests that low expression intensity may have resulted in inaccurate expression level determinations by 1 or both arrays. Poor correlations between 2 different technologies for genes that are expressed at low levels have also been reported by others and suggest that measurements for low expression genes are unreliable.[15-19] The different Caco-2 clones used in the 2 studies may also explain the lack of agreement in expression changes upon differentiation. Moreover, different probe domains could result in different expression levels if alternative mRNA splicing occurs in the gene product

being measured. However, in order to ascertain which of the 2 arrays accurately determines the up- or downregulation of the disputed genes, a third reliable source of gene expression changes following Caco-2 cell differentiation would be necessary. Quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) analysis on a few genes included in Table 3 has been reported previously.[8] The RT-PCR results confirm that GLUT5 and ABC1 are strongly upregulated (6.91- and 5.01-fold, respectively) and are consistent with the Affymetrix array data for GLUT5 expression changes and with the custom microarray data for changes in ABC1 expression. Both arrays reported ABCC3 upregulation, which was also confirmed by previous RT-PCR results.[8]

Normalization of the acquired raw data is often employed to improve the quality and accuracy of differential expression by filtering most of the nonbiological variation associated with slide/chip effects, background noise, and systematic errors.[39] The linear normalization used in the Affymetrix analysis system assumes the data are normally distributed and are valid for a large data set. However, systemic biases are not always linear, and nonlinear normalization methods such as Lowess normalization have been used to control signal-dependent nonlinear bias between Cy5 and Cy3 channels in spotted arrays.[22,25,40] Linear normalization shifts the data toward the center of the distribution toward zero but does not affect the data spread. In contrast, the nonlinear Lowess method reduces the spread compared with the globalization or linear method.[22] Consequently, the Affymetrix data generated $\log_2$ ratios ranging from +2.59 to -2.24, while the spotted array expression differences were in the range of +1.15 to -1.20.

Conclusion

The gene expression results from the 2 different arrays suggest that the data generated with Affymetrix arrays and a custom-designed oligonucleotide array can be comparable, even though there are differences in technology and experimental design. In this study, the number of gene probes available for comparison between the 2 array systems was limited; however, both arrays detected similar changes in transporter and ion channel gene expression upon Caco-2 cell differentiation. The observed correlations between the 2 sets of array data are especially good when the genes being compared are expressed abundantly. The poor correlation between the 2 arrays for other genes may be attributed to low expression levels and to differences in probe-related factors such as sequence length, sequence overlap, and cross-hybridization in the 2 arrays. The comparisons described here reflect a common practice of different Caco-2 cells being used by researchers in various laboratories; the results validate some of the conclusions drawn in the 2 independent studies on mRNA profiles and changes upon Caco-2 differentiation.

Acknowledgements

This research was supported by NIH Grant GM37188 and in part by NIH General Medical Sciences Grant GM61390, Plasma Membrane Transporter Group at UCSF. We thank Thomas Pastijn for Access programming support.

References

1. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.

2. Hughes TR, Marton MJ, Jones AR, et al. Functional discovery via a compendium of expression profiles. *Cell*. 2000;102:109-126.

3. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature*. 2000;405:827-836.

4. Sun D, Lennernas H, Welage LS, et al. Comparison of human duodenum and Caco-2 gene expression profiles for 12,000 gene sequences tags and correlation with permeability of 26 drugs. *Pharm Res*. 2002;19:1400-1416.

5. Chay CH, Cooper CR, Gendernalik JD, et al. A functional thrombin receptor (PAR1) is expressed on bone-derived prostate cancer cell lines. *Urology*. 2002;60:760-765.

6. Landowski CP, Sun D, Foster DR, et al. Gene expression in the human intestine and correlation with oral valacyclovir pharmacokinetic parameters. *J Pharmacol Exp Ther*. 2003;306:778-786.

7. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*. 1996;14:1675-1680.

8. Anderle P, Rakhmanova V, Woodford K, Zerangue Z, Sadee W. Messenger RNA expression of transporter and ion channel genes in undifferentiated and differentiated Caco-2 cells compared to human intestine. *Pharm Res*. 2003;20:3-15.

9. Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res*. 2000;28:4552-4557.

10. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270:467-470.

11. Anderle P, Duval M, Draghici S, Kuklin A, Littlejohn TG, Medrano JF, Vilanova D, Roberts MA. Gene expression databases and data mining. *Biotechniques*. 2003;Mar(suppl):36-44.

12. Ishii M, Hashimoto S, Tsutsumi S, et al. Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics*. 2000;68:136-143.

13. Bittner M. Obtaining and evaluating gene expression profiles with cDNA microarrays. In: Suhai S, ed. *Genomics and Proteomics*. New York, NY: Kluwer; 2000:5-25.

14. Li J, Pankratz M, Johnson JA. Differential gene expression patterns revealed by oligonucleotide versus long cDAN arrays. *Toxicol Sci*. 2002;69:383-390.

15. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from 2 different microarray technologies. *Bioinformatics*. 2002;18:405-412.

16. Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. Accuracy and calibration of commercial oligonucleotide and custom cDNA arrays. *Nucleic Acids Res*. 2002;30:e48.

17. Tan PK, Downey TJ, Spitznagel EL, et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*. 2003;31:5676-5684.

18. Barczak A, Rodriguez MW, Hanspers K, et al. Spotted long oligonucleotide arrays for human gene expression analysis. *Genome Res*. 2003;13:1775-1785.

19. Yang MC, Ruan QG, Yang JJ, et al. A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol Genomics*. 2001;7:45-53.

20. Shin HC, Landowski CP, Amidon GL. Transporters in the GI tract. In: Waterbeemd V-d, Lennernäs H, Artursson P, eds. *Drug Bioavailability/Estimation of Solubility, Permeability and Absorption*. Weinheim Germany: Wiley-VCH; 2003.

21. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997;278:680-686.

22. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple systematic variation. *Nucleic Acids Res*. 2002;30:e15.

23. Mariadason JM, Corner GA, Augenlicht LH. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin a, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer Res*. 2000;60:4561-4572.

24. Mariadason JM, Arango D, Corner GA, et al. A gene expression profile that defines colon cell maturation *in vitro*. *Cancer Res*. 2002;62:4791-4804.

25. Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, Quackenbush J. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*. 2002, 3(11):research0062.1-0062.12.

26. Pedron T, Thibault C, Sansonetti PJ. The invasive phenotype of *Shigella flexneri* directs a distinct gene expression pattern in the human intestinal epithelial cell line Caco-2. *J Biol Chem*. 2003;278:33878-33886.

27. Hill AA, Brown EL, Whitley MZ, Tucker-Kellogg G, Hunter CP, Slonim DK. Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biol*. 2001;2(12):research0055.

28. Zarrinkar PP, Mainquist JK, Zamora M, et al. Arrays of arrays for high-throughput gene expression profiling. *Genome Res*. 2001;11:1256-1261.

29. Baum M, Bielau S, Rittner N, et al. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res*. 2003;31:e151.

30. Engle MJ, Goetz GS, Alpers DH. Caco-2 cells express a combination of colonocyte and enterocyte phenotypes. *J Cell Physiol*. 1998;174:362-369.

31. Levy P, Robert A, Picard J. Biosynthesis of glycosaminoglycans in the human colonic tumor cell line Caco-2: structural changes occurring with the morphological differentiation of the cells. *Biol Cell*. 1988;62:255-264.

32. Hidalgo IJ, Raub TJ, Borchardt RT. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology*. 1989;96:736-749.

33. Rothen-Rutishauser B, Braun A, Günthert M, Wunderli-Allenspach H. Formation of multilayers in the Caco-2 cell culture model: a confocal laser scanning microscopy study. *Pharm Res*. 2000;17:460-465.

34. Mariadason JM, Kurt L. Rickard KL, David H, Barkla DH, Augenlicht LH, Gibson PR. Divergent phenotypic patterns and commitment to apoptosis of Caco-2 cells during spontaneous and butyrate-induced differentiation. *J Cell Physiol*. 2000;183:347-354.

35. Fleet JC, Wang L, Vitek O, Craig BA, Edenberg HJ. Gene expression profiling of Caco-2 BBe cells suggests a role for specific signaling pathways during intestinal differentiation. *Physiol Genomics*. 2003;13:57-68.

36. Behrens I, Kissel T. Do cell culture conditions influence the carrier-mediated transport of peptides in Caco-2 cell monolayers? *Eur J Pharm Sci*. 2003;19:433-442.

37. Bravo SA, Nielsen CU, Amstrup J, Frokjaer S, Brodin B. In-depth evaluation of Gly-Sar transport parameters as a function of culture time in the Caco-2 cell model. *Eur J Pharm Sci*. 2004;21:77-86.

38. Kothapalli R, Yoder SJ, Mane S, Loughran TP Jr. Microarray results: how accurate are they? *BMC Bioinformatics*. 2002;3:22.

39. Geller SC, Gregg JP, Hagerman P, Rocke DM. Transformation and normalization of oligonucleotide microarray data. *Bioinformatics*. 2003;19:1817-1823.

40. Tseng GC, Oh M-K, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*. 2001;15:2549-2557.